Speech Processing 2016/2017

2nd Test

April 28th 2017

Please identify this form with your name and student number in the reserved spaces at the beginning and end of the test. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers to these questions will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

Name:	
Number:	
Group number:	

1. Classify as True (T) or False (F) on the left side of each item.

   (a) The very low bit rate coders typically suffer from intelligibility and speaker recognizability problems.

   (b) The only difference between LPC vocoders and MELP is in the jitter simulation.

   (c) In standardized RPE coders, the equispaced pulse grid models the harmonicity in voiced speech.

   (d) In subband coders, the high frequency subbands are quantized with a higher number of bits per sample than the low frequency subbands.

   (e) The robustness to channel errors is the main disadvantage of using differential encoding schemes.

   (f) Articulatory models can be used for very low bit rate coding because articulators vary slowly.

   (g) Prosody varies across languages but not across varieties of the same language.

   (h) In concatenative synthesis systems, speech segments are always concatenated without separating the source and the vocal tract contributions.

   (i) European Portuguese has more than 30 different visemes.

   (j) CART based approaches for prosodic phrasing may place breaks between function and content words.

2. Which types of delay should be considered in speech coding?

   _____
   _____
   _____
   _____
   _____
   _____

3. Consider a coder that uses both pitch prediction and LPC prediction.

   (a) Write the Z-transform of the transfer function of both predictors (Pd and Ps, respectively).

(b) Specify the bit allocation for a coder using these two predictors, such that the (sub)total bit rate for both is $<=$ 2.6kpbs. Use the table below.

Parameter	Updating	No. Bits / Frame	Sub-total
Pitch predictor	every		
LPC predictor	every		

(c) Name at least three coders that could use these two predictors.

_____

_____

_____

(d) If the excitation for the cascade of two predictors is encoded using a narrowband waveform quantization scheme of only 2 bits per sample, what would be the total bit rate for this hypothetical coder?

_____

_____

(e) If a noise shaping block is integrated in this coding scheme, will this block contribute to increase the signal-to-noise ratio of this coder?

_____

4. Indicate which are the waveform generation techniques that may be considered state of the art in TTS systems in the decades shown below:

70s-80s	
80s-90s	
90s-00s	
00s-10s	
10s-20s	

5. The implementation of a formant synthesizer done in Lab 3 was used to synthesize a neutral vowel, with fixed F0=100 Hz, and Fs=8000 Hz.

(a) Compute the angles (in degrees) of all the possible formants, and draw the corresponding poles in the complex plane, assuming a magnitude of 0.95, using the unit circle as a reference. (Please indicate how you computed the angles.)

(b) By how much would you have to reduce the bandwidth of the formants in order for the change to be perceptually noticeable?

_____

Indicate the main changes in the previous figure using arrows, without computing values.

6. Which break index in the ToBI model indicates maximum coarticulation between consecutive words?

_____

Which one indicates maximum disjunction?

_____

7. Write the nativised broad phonetic transcription of the following sentence, using only EP phones, as you would hear it in the Portuguese media.

Follow your Twitter and Facebook friends.

```



```

8. Write the phonetic transcription of the acronyms BASF, LUSOSIDER and PSP, and indicate for each one if the corresponding transcription should be included in a lexicon of exceptions (LEX) or could be derived by rule (RULE).

_____
_____
_____

9. Consider the two isolated words *apagador, pensamento*. Indicate which phone sequences from these words would you select to concatenate each word in the following list. Write NOT POSSIBLE if the word could not be synthesized using the typical process in concatenative speech synthesis.

apaga	
apagamento	
mentor	
paga	
pagador	
pagamento	
pensa	
pensador	
penso	

10. The prosody generation module computed the target value for the duration of vowel *6*: 80 ms. The left sub-unit is *s6* with duration values of 44 and 36 ms for the two parts. The right subunit is *6d* with duration values of 54 and 22 ms for the two parts. Indicate the target durations for the two sub-units of the vowel (just indicate the computation).

The prosody generation module also computed the target value for the pitch period of the vowel *6*: 5 ms. The original values of the left and right subunits are 6 and 4 ms, respectively. Which prosody modification should be applied first?

_____

11. Using the following syntax

a → b / m .. n

describe the rules for grapheme *c*, where m and n may be graphemes (e.g.: a, b, etc.), classes of graphemes (vowels, consonants, etc.), the word boundary (#), or any grapheme (*). You may use the symbol *0* to mark phonemic nulls, and the symbols | and ( ) to mark disjunction between several graphemes (e.g.: a | b | c). The rules are applied in order, until one matches the context and, in this case, the following rules are not applied. The rules do not need to contemplate compound words, or words of foreign origin, but they must follow the orthographic agreement.

1	3.0
2	1.0
3	4.0
4	1.5
5	2.0
6	1.0
7	1.0
8	1.2
9	1.8
10	1.5
11	2.0